DOCUMENT RESUME

ED 456 274 CE 082 254

AUTHOR Evans, Elizabeth A.
TITLE Interpreting Educational Research Using Statistical Software.
PUB DATE 2001-03-00
NOTE 11p.; In: Research to Reality: Putting VET Research To Work. Proceedings of the Australian Vocational Education and Training Research Association (AVETRA) Conference (4th, Adelaide, Australia, March 28-30, 2001); see CE 082 232.
AVAILABLE FROM For full text: http://www.avetra.org.au/PAPERS%202001/evans.pdf.
PUB TYPE Numerical/Quantitative Data (110) -- Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Oriented Programs; *Computer Software; Computer Uses in Education; Cost Effectiveness; *Data Interpretation; Developed Nations; *Educational Research; Foreign Countries; Literacy Education; Postsecondary Education; *Research Methodology; *Statistical Analysis; Tutoring; Vocational Education
IDENTIFIERS Statistical Package for the Social Sciences

ABSTRACT

A live demonstration of how a typical set of educational data can be examined using quantitative statistical software was conducted. The topic of tutorial support was chosen. Setting up a hypothetical research scenario, the researcher created 300 cases from random data generation adjusted to correct obvious error. Each case represented a student who was required to take a literacy test; depending test results, the student was recommended to access a quantity of tutorial support hours. Use of tutorial hours was recorded. At the end of the semester, each took a communications test. Eleven variables were identified. The quantitative statistical analysis software used was SPSS (Statistical Package for the Social Sciences) Version 10. To get a sense of the literacy test marks and factors that impacted on marks achieved, the researcher ran "Explore" to get a table to show the shape of the distribution of marks by faculty and got a histogram and boxplot. An independent samples t-test was run to obtain a probability statement about the difference in means between genders. A one-way ANOVA (analysis of variance) showed the significance of means among faculties. SPSS calculated the Pearson Correlation on variables to produce tables that quantify the strength of the relationship between variables. The SPSS-prepared Matrix Scatterplot was used in regression analysis to predict a variable from one or more predictor variables. (YLB)

Reproductions supplied by EDRS are the best that can be made
from the original document.

Interpreting educational research using statistical software

Elizabeth A Evans

University of Sydney, Cumberland Campus, New South Wales

ED 456 274

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

K. Whittingham

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

The VET sector naturally produces a gold mine of data waiting to be retrieved, assimilated and interpreted into meaningful information. Senior staff need this information to make informed decisions to improve student outcomes and optimise scarce resources. There are a number of software products on the market that will assist VET staff and researchers in their quest for answers to complex questions. Providing the right question is asked (and even if it is not the right question to ask), the use of these software products can save hours of drudgery and the user does not have to be a statistician to find the answers.

This paper is based on a live demonstration of how a typical set of educational data can be examined using quantitative statistical software. A research question will be investigated by examining the data to establish if a relationship exists between two or more variables, such as hours of tutorial support and resulting grades.

It is not my intention to discuss the academic governance of research protocols. In fact, I debated as to whether or not I should write a paper, given that I was demonstrating the benefits of a computer software product to do quantitative statistical analysis of research data. Then I thought that I should, so that the first impression I expected from the audience would be reinforced, encouraging them to think about how useful statistical analysis could be in making timely data-based decisions.

When I submitted my abstract, I was working for the Technical and Further Education Institute (TAFE) of New South Wales (NSW) and intended to use some 'real' VET data. Between then and now, I've moved onto the University of Sydney and so I have lost that opportunity, although I would not consider it to be vitally important. It's not the data itself, but the statistical analysis of it that I wanted to show you.

I decided to pick the topic of tutorial support because it has been widely debated within the VET sector, both in terms of its cost and its benefit. Setting up a research scenario in my mind, I created 300 cases from random data generation which I then adjusted to correct obvious error (for example, you could not be born overseas, be 22 years old, and have been in Australia for 30 years), and then further to be able to highlight the analysis.

Each case represents a student. Each student was required to take a literacy test (for both numeracy and language), and from those results the student was recommended to access a quantity of tutorial support hours. Their use of tutorial hours was recorded, and at the end of the semester, all of them took a communications test.

CE022254

BEST COPY AVAILABLE

There were 11 variables identified: Gender; Age; Country of origin – collated into 5 global regions; Native language – first language; Years in Australia; Prior Education - < Grade 10, Grade 10, HSC, Tertiary; Faculty enrolled – six different faculties; Initial literacy test mark, Tutorial hours required, Tutorial hours used; and Final communications exam mark.

The quantitative statistical analysis software I used was SPSS Version 10. There are other similar products on the market, and it is worthwhile comparing them. I learned this software through compulsory research courses in graduate school, and considered it a valuable tool for management to use in a practical way to explore and measure their environment, and their interaction with it.

Research is all about getting to know your data. SPSS allows you to look at your data from all sorts of angles, and this is known as 'descriptive' statistics.

As a layperson, I would like to know how many students are female or male, and I would like to know how many students are in each faculty. I can find this out by using Frequencies from the Analyse – Descriptive Statistics menu.

GENDER

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	131	43.7	43.7	43.7
	Male	169	56.3	56.3	100.0
	Total	300	100.0	100.0	

FACULTY

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	General Education	77	25.7	25.7	25.7
	Business	63	21.0	21.0	46.7
	Tourism & Hospitality	55	18.3	18.3	65.0
	ITAM	40	13.3	13.3	78.3
	Engineering & Manufacturing	37	12.3	12.3	90.7
	Rural & Mining	28	9.3	9.3	100.0
	Total	300	100.0	100.0	

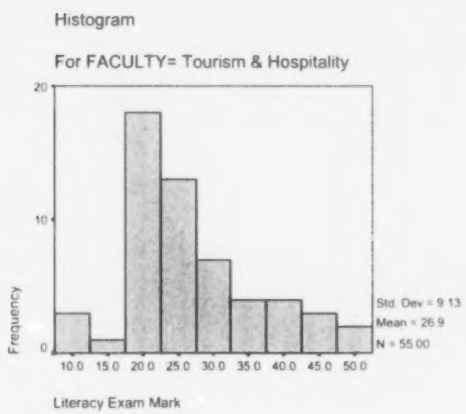
This gives me some valuable information. It confirms that I have 300 cases and no 'missing' cases, and it gives me the percentage contribution of each category of the variable.

What I'm really after is some sense of the literacy test marks, and there could be many things that could impact upon the marks achieved, such as prior education, country of origin, what a student's first language is, and how long that student has

been in Australia. But first I would like to see if the literacy marks are different from one faculty to another.

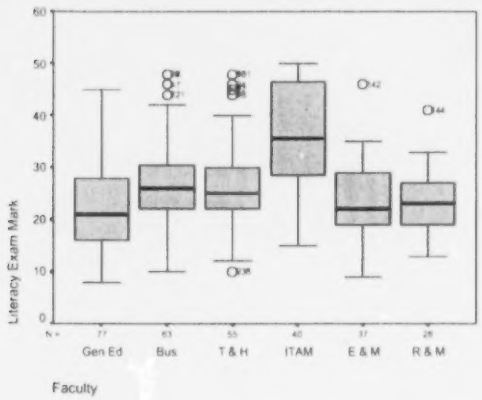
By running 'Explore', I will get a table that shows me the shape of the distribution of the marks in each of the faculties. Shape is defined by a number of statistics. What is the average literacy mark in General Education or in Engineering and Manufacturing? How far do the marks range? Are the marks 'concentrated' around the average or are they 'spread out'? SPSS provides very comprehensive descriptive statistics under Explore, but what is even more 'enlightening' to the layperson is the graphical representations.

Below is a histogram, or frequency chart of the literacy test marks for the faculty of Tourism and Hospitality. I could have asked for a 'normal curve' to be included, which would enhance the imagery of the spread of the data. You can see that the majority of the marks fell +/- 20 marks out of a possible 50 marks. SPSS provides some useful information such as the average mark (26.9), the standard deviation (26.9 +/- 9.13) and the number of cases (students) who took the literacy test and are enrolled in this faculty.



SPSS wraps up its 'Explore' analysis by giving you what is called a boxplot.

This gives you a helicopter view of the literacy marks across all faculties. The box itself represents the interquartile range of the data (middle 50%), while the 'whiskers' are the last data values within 1.5 lengths of the box. The heavy line within the box is the median, and the little circles that lie outside the whiskers are 'outliers', while the appearance of asterisks would be 'extremes'. This is valuable to know, because it means the data value is 'unusual', and you may want to check that the data entered was correct and not a mistake.



We may also want to explore if there is a difference between females and males. Using the literacy test mark as our focus, we can run an Independent samples t-test to obtain a probability statement about the difference in means between females and males. The following table is produced:

Independent samples test

		Levene's Test fo Equality o Variances		t-test fo Equality o Means		
		F	Sig.	t	d	Sig. (2-tailed)
LITERACY Literacy Exam Mark	Equal variances assumed	.094	.759	.594	298	.553
	Equal variances no assumed			.597	284.409	.551

There are 131 females and 169 males, so the groups are not homogeneous (as is often the case). The t-test assumes homogeneity, so SPSS gives you two sets of t-test results. For the means to be significantly different between females and males, the Levene's Test would have to be less than .05 – but this is not the case. The significance of the t-test is much greater than .05, so there is no significant difference between the literacy test scores for males and females.

Do the mean literacy scores between faculties differ? Sometimes it is obvious from the boxplot. In this case, what we can tell here is that ITAM (Information Technology Arts & Media) has a higher range of marks than all other faculties and that its median literacy test mark looks to be significantly higher than the rest. We can ask SPSS to Compare Means, and SPSS gives you a few options, from comparing the actual means into a simple table, to conducting various t-Tests (for two populations), and performing a one way ANOVA (ANalysis Of VAriance for two or

more populations). A one way ANOVA shows you how significantly the mean of a given faculty differs to means of the other faculties.

ANOVA

LITERACY

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6176.682	5	1235.336	16.142	.000
Within Groups	22499.088	294	76.528		
Total	28675.770	299			

This is the first table produced. We are interested in the mean differences within the groups, not just between them. In this paper, I have assumed, as is commonly the practice in social research, that the level of significance is .05; anything below is significant, and anything above is not. The 'Sig' here is .000, and that means there is virtually no probability of us achieving sample means that are different by chance alone. Therefore, the means between the faculties are different, but which ones?

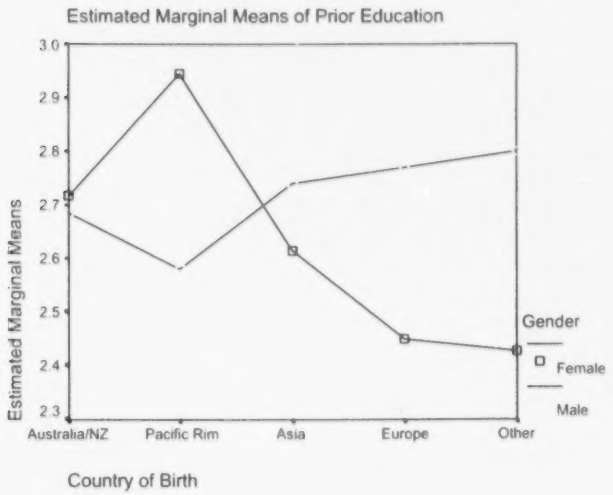
(I) FACULTY	(J) FACULTY	Mean difference (I-J)	Std. error	Sig.
General Education	Business	-5.94*	1.49	.001
	Tourism & Hospitality	-4.99*	1.54	.021
	ITAM	-14.30*	1.71	.000
	Engineering & Manufacturing	-1.13	1.75	1.000
	Rural & Mining	-1.72	1.93	1.000
Business	General Education	5.94*	1.49	.001
	Tourism & Hospitality	.95	1.61	1.000
	ITAM	-8.37*	1.77	.000
	Engineering & Manufacturing	4.80	1.81	.127
	Rural & Mining	4.21	1.99	.521
ITAM	General Education	14.30*	1.71	.000
	Business	8.37*	1.77	.000
	Tourism & Hospitality	9.32*	1.82	.000
	Engineering & Manufacturing	13.17*	2.00	.000
	Rural & Mining	12.58*	2.16	.000

* The mean difference is significant at the .05 level.

This table shows statistically how significantly different each of the faculty means on the left-hand column is to each of its counterparts. I've culled the faculties of Tourism & Hospitality, Engineering and Manufacturing and Rural & Mining from the imported table to keep it brief, but note that, as an example, the Faculty of Business' mean literacy score is significantly different from General Education and ITAM, but not from T&H, E&M and R&M. As expected, ITAM's mean literacy score is significantly different from all of its counterparts. Again, this is useful information, and sometimes the eyes on the boxplot can be deceived!

When we move beyond one-way or one-factor ANOVA, the distinction between main effects and interactions become relevant. A main effect is an effect (or group difference) due to a single factor (independent variable). For example, we want to study the prior education difference across country of origin and by gender. The effect of country of origin alone, and the effect of gender alone, would each be considered a main effect. In other words, we want to test that the country of origin differences are identical for each gender group. Since we are studying two factors, there can only be one interaction. Sound clear as mud? Let us look at the graph.

Visually the overall means for men and women are not the same (women are higher). What is important is that the gender differences vary dramatically over the countries of origin: in Australia, it's about the same, but in the Pacific Rim, women have a higher education than men, while in Asia they have a lower level of education, and this is lower still in Europe. From the table below, there are no main effects and a strong interaction, but the 'model' accounts for only 2% (R^2=.019) of the variance in education. There must be other factors that cause the mean differences.



Tests of Between-Subjects Effects

Dependent Variable: PRIORED Prior Education

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4.635[a]	9	.515	.382	.943
Intercept	1777.255	1	1777.255	1316.983	.000
ORIGIN	.736	4	.184	.136	.969
GENDER	.440	1	.440	.326	.569
ORIGIN * GENDER	3.934	4	.983	.729	.573
Error	391.352	290	1.349		
Total	2540.000	300			
Corrected Total	395.987	299			

a. R Squared = .012 (Adjusted R Squared = -.019)

Correlation is about quantifying the strength of the relationship between variables. The Pearson product-moment correlation coefficient is a measure of the extent to which there is a linear or straight line relationship between two variables and the value will fall between -1 (perfectly negative correlation - as one moves up, the other moves down) to +1 (perfectly positive correlation - for every unit move in one, there is an identical move in the other). SPSS will calculate the Pearson Correlation on any number of variables you choose and produce a table like the one below. It is easy to interpret. As I mentioned at the beginning, I manipulated the randomly generated data to enhance the SPSS displays. In this case there is a significant correlation of each variable to all others selected, as indicated by the **. Some of them are positive correlations, ie the higher the prior education the higher the literacy score, and some of them are negative; the higher the literacy score, the less tutorial hours required. 'Real' data usually is not so accommodating. In social or market research where straight-line relationships are found, significant correlation values are often between .3 and .6. But what if the relationship is non-linear? SPSS accommodates this with curve estimates.

Correlations

		PRIORED	LITERACY	TUTREQD	TUTUSED	COMMSC
PRIORED	Pearson Correlation	1.000	.552**	-.602**	-.448**	.469**
	Sig. (2-tailed)	.	.000	.000	.000	.000
	N	300	300	300	300	300
LITERACY	Pearson Correlation	.552**	1.000	-.868**	-.477**	.908**
	Sig. (2-tailed)	.000	.	.000	.000	.000
	N	300	300	300	300	300
TUTREQD	Pearson Correlation	-.602**	-.868**	1.000	.559**	-.778**
	Sig. (2-tailed)	.000	.000	.	.000	.000
	N	300	300	300	300	300
TUTUSED	Pearson Correlation	-.448**	-.477**	.559**	1.000	-.186**
	Sig. (2-tailed)	.000	.000	.000	.	.001
	N	300	300	300	300	300
COMMSC	Pearson Correlation	.469**	.908**	-.778**	-.186**	1.000
	Sig. (2-tailed)	.000	.000	.000	.001	.
	N	300	300	300	300	300

**. Correlation is significant at the 0.01 level (2-tailed).

BEST COPY AVAILABLE

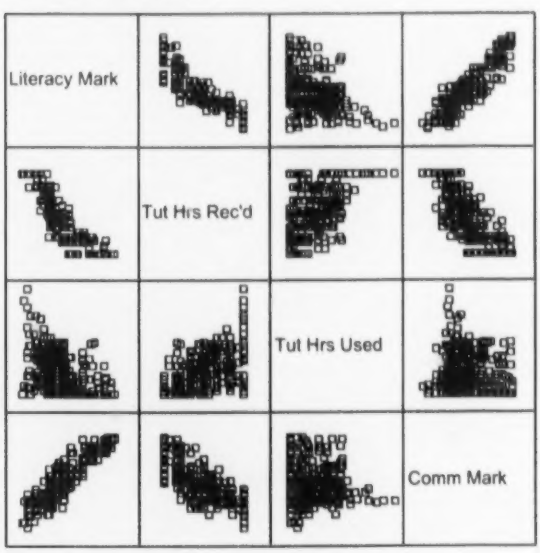
Correlation analysis provides a neat single numeric summary of the relationship between two variables, but it would be more useful from a practical point of view to have some form of predictive equation. Regression analysis is a statistical method used to predict a variable from one or more predictor variables. The first thing we do is construct a scatterplot to get a vision of the relationship.

The scatterplot SPSS has prepared is called a Matrix Scatterplot, and what it shows is a scatterplot for each variable to all other variables. As you would expect, there is likely to be a positive linear relationship in some, a negative one in others, and little or no relationship in yet some others.

Visually, it would appear that there is a positive relationship between one's literacy mark and one's final communication mark; if you did well in the initial literacy test, it is likely that you did well in the final communication test. There appears to be a negative relationship between the literacy mark and the tutorial hours recommended. Again, this makes sense - if you scored well on the literacy test, the tutorial hours recommended would be low or nil.

There is also a negative relationship between Tutorial Hours Recommended and the Communication Mark (the logic of which follows from the literacy test). There appears to be little or no relationship between the literacy mark and the tutorial hours used, nor between the tutorial hours recommended and the tutorial hours used, nor between the tutorial hours used and the communication mark. It may be that the tutorial hours do not have much effect on the outcome!

Matrix Scatterplot



BEST COPY AVAILABLE

Because we have already discovered through correlation that the variables were related, we have bypassed simple regression for multiple regression, where we will construct a predictive model to estimate the communication mark to be expected if there are certain values given to the literacy mark, the tutorial hours recommended, and those used.

The important number in the model summary is the R Square. The R Square measure of .911 indicates that these three predictor variables account for about 91.1% of the variation in the final communications mark. In real life, it is not likely that you get this close!

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.954[a]	.911	.910	2.79

a. Predictors: (Constant), TUTUSED Tutorial Hours Used, LITERACY Literacy Exam Mark, TUTREQD Tutorial Hours Recommended

b. Dependent Variable: COMMSC Communications Mark

Coefficients

Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.
		B	Std. Error	Beta		
	(Constant)	5.152	1.289		3.996	.000
	LITERACY Literacy Exam Mark	.885	.033	.935	26.736	.000
	TUTREQD Tutorial Hours Recommended	-.158	.036	-.163	-4.398	.000
	TUTUSED Tutorial Hours Used	.311	.019	.352	16.778	.000

Dependent Variable: Communications Mark

The coefficient table produced by SPSS gives you the Betas you need for your multiple regression model of: Y = (ß1 * X1)+(ß2*X2)+(ß3*X3)... What is really important here is that you cannot predict outside of your range of existing values, so you must pick a literacy mark between the lowest and highest recorded (X1), and the same applies for the tutorial hours recommended (X2) and the tutorial hours used (X3).

What is important to note here is that at the end of the day, in these cases, the final communication mark principally depends on the mark achieved in the initial literacy test. If the cases were 'real' (and I acknowledge that they are not, and perhaps for the best), this would give administrators some food for thought about the resources they devote to tutorial support as it relates to educational outcomes.

Conclusion

I would like to be able to tell you just how easy all of this is, and how the software can be used in a variety of ways to the benefit of your institution. I can certainly attest to the latter, but like all software, your proficiency is a matter of time and effort. An introductory statistics course (for Dummies) would not go astray, as it has a language of its own, although the concepts are quite intuitive.

You should also know that I am an administrator, and not a professional (or otherwise) researcher, so I am always looking for information and I am always concerned with expenditure! If you researchers out there thought this paper was not to academic standards, then you are absolutely right. It has been written for the lower earthlings who struggle for answers every day.
One thing that you may have noted is that I have imported the SPSS output into this word document alternatively as an object, and also as a 'copy/paste'. SPSS (colourful) output can really enhance your reports.

Where this kind of software is a powerful tool, and can save you a significant amount of money, is in market research. Educational institutions have realised how important it is to measure their environment and their customers. Without debating the issues or reliability and validity (ie having someone externally prepare the measurement tool, administer and analyse the results), it can be very useful to analyse secondary data (such as the ABS) as well as primary data (conducting your own student satisfaction survey).

I encourage you to explore these products and invest a little to get a lot. Happy analysing!

Contact details

Elizabeth Evans
Email: E.Evans@cchs.usyd.edu.au